图书情报工作
第64卷 第1期 2020年1月

ChinaXiv合作期刊

中外情报学论文创新性特征研究*

■ 曹树金 闫欣阳 张倩 卓伊玲
中山大学资讯管理学院 广州 510006

摘 要：[目的/意义]综合运用定性与定量相结合的方法对近年中外情报学论文的创新性进行分析和对比，揭示情报学领域研究的创新性特征，发现领域学术论文中创新句内部的知识关系，进行更细粒度的论文创新性分析，为研究领域创新点深层次利用提供条件，同时丰富科技论文创新性监测的途径，促进科学研究创新。[方法/过程]从句子级创新性识别出发，选取中英文各两种情报学期刊作为样本，采用信息抽取和机器学习的方法，将创新句的抽取从现有的摘要扩展到全文，充分利用句子结构和句法特征识别领域创新内容，探讨近年中外情报学论文在创新对象、主题、类别等方面的特征，并做对比分析，最后通过对自动分类的论文集合进行定性的内容分析，总结归纳出中外情报学论文创新的表达范式。[结果/结论]从创新的表达来看，中外情报学论文创新句的分布情况基本一致，英文期刊论文创新的表达更丰富。从创新性特征来看，英文情报学期刊论文创新主题较集中，而中文主题多样和分散；具体方法的创新是近年情报学领域的创新热点，而在研究方法上创新不足；中英文情报学期刊论文的创新性特点都反映了应用研究、实证研究的成果较多，而理论创新推动缓慢的趋势。

关键词：创新性特征 学术论文 句法解析 句子分类
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.01.011

chinaXiv:202304.00380v1

1 引言

创新是引领发展的第一动力，是科学研究的核心工作，是学术论文的本质要求。R. Tan 曾提到科学研究创新性："一个人的研究价值不在于他投入多少努力，相反，研究的价值在于结果的创新性"[1]。创新性是科技论文学术质量评价的重要标准，是决定其学术水平的核心、关键特质及发表与否的首要依据[2]。大数据时代，对于研究者来说，从文献中快速准确地获取创新观点是亟待解决的需求，从学科的宏观角度来看，监测创新、促进创新也是科学研究发展的本质要求。而创新本身及其表述的复杂性和多样性又为识别创新特征增添了难度。已有研究从事件[3-4]、文档[5-6]和句子[7-9]级别对创新特征的自动识别进行了探索，但在实际应用中，学术论文创新性的评价，仍以费时费力且主观性较强的同行评议为主要手段，针对学术论文的创新性研究仍具有较大的发展空间[10]。

抽取领域学术论文集中的创新点进行分析，能有效揭示领域的创新进展以及创新点的类型及影响[10]。本研究拟对情报学论文的创新性特征进行对比探究，从句子级创新性识别出发，充分利用句子结构和句法特征，采用信息抽取和机器学习的方法分析创新的方面、对象及主题，由广而精衡量情报学研究近年的创新特征与进展。本研究的创新之处在于，将创新句的抽取从现有的摘要扩展到全文，而不仅停留在摘要集层面，并在信息抽取和机器学习的方法应用中充分融入句子结构信息。本研究的目的不仅在于揭示某一领域的创新情况，更重要的是发现领域学术论文中创新句内部的知识关系，为研究领域创新点的深层次利用提供途径，为细粒度知识组织和检索奠定基础，助力大数据环境下知识的推理和发现，同时丰富科技论文创新性监测的途径，最终达到促进科学研究创新的目的。

2 文献综述

2.1 学术论文创新性的涵义与表现

对于"创新性"，很难给出一个精准的定义。《韦

* 本文系国家社会科学基金重大项目"基于特定领域的网络资源知识组织与导航机制研究"（项目编号：12&ZD222）研究成果之一。
作者简介：曹树金（ORCID：0000-0003-1855-4522），教授，博士生导师，E-mail：caosj@mail.sysu.edu.cn；闫欣阳（ORCID：0000-0001-8436-1927），硕士研究生；张倩（ORCID：0000-0002-6390-6112），硕士研究生；卓伊玲（ORCID：0000-0001-7522-6328），硕士研究生。
收稿日期：2019-12-03 本文起止页码：80-92 本文责任编辑：杜杏叶

氏词典》将“novelty”定义为“新的或与任何熟悉事物不同的事物”[11]。创新信息(novel information)是指包含新内容的句子,通常被定义为冗余的反义词[12]。在 J. Allan 等[13]的研究中,创新性被描述为基于句子中存在的新单词的新信息。B. Uzzi 等[14]认为,科学创新是通过激发新见解的原创组合产生的。从知识组合的角度来看,创新性可以被定义为以前所未有的方式重新整合已有的知识,这种组合的观点也被各学科的学者所接受[15]。这些表述的核心内涵基本一致,主要在于用“新”方式组合已有的或新的知识。关于学术论文的创新性,周露阳[16]认为有三层基本含义:①与已有学术文献“不同”的论文。这种不同可以是对已有文献的“局部改进”,也可以是与已有文献完全不同的“全新”。②这种不同是“所涉学术领域之知识或信息”的不同。③这种不同的知识或信息须是“有价值”的。本研究所进行的学术论文创新性识别,即对论文作者在文中表述的研究创新之处进行挖掘和特征分析。

2.2 学术论文创新性评价

长期以来,人文社科期刊论文的学术创新性评价方法主要包括两种:以同行评价为主要手段的定性评价法和基于文献计量学的定量评价法。学术论文创新性评价的文献计量法包括单个特征指标评价法、以影响力测度创新力法、指标体系评价法、基于论文内容的评价法等[17]。同行评议这种评价方法虽然被广泛使用,但在评议的过程中存在利益冲突、主观性强、操作空间大、耗时长、效率低下等弊端。刘丽萍和刘春丽[18]指出同行评议的弊端主要有:该方法增加了编辑和审稿人之间的利益冲突,公众评议观点受质疑,审稿人刻意评议,作者违心修改论文等问题。

一些学者使用论文的单个特征指标作为论文学术创新性评价指标,比如以论文作者的 h 指数、论文的被引量、论文参考文献的影响力或期刊的影响因子等单个指标为主的评价方法[19]。Y. Lee 等[20]通过计算参考文献每一期刊对的相似程度,将论文所有期刊对的相似程度从小到大排列并取前 10% 分位数,再取负数,所得数值越大,该论文的新颖性越强。L. Wu 等[21]在 NATURE 上提出了一个新的创新测度指标 Disruption,通过划分引文结构量化论文新颖性。也有学者综合运用社会网络分析法和统计学方法分析,通过相关分析、回归分析以及结构方程模型等验证各种指标与文献创新力之间的关系[22],叶继元[23]从社会、经济、文化、政治等角度解析了我国学术期刊论文质量与创新评价体系存在的弊端和原因,提出了形式评价、内容评价和效用评价的“三位一体”新概念组合,对我国社科期刊论文质量评估和创新测度方面的研究有较强的理论指导意义。

综上,很多学者用统计学方法对作者声望、期刊影响力、被引次数、下载次数与论文创新度的关系进行了实证分析,得出的结论各有不同,但基本上都认为仅以这些指标去衡量论文创新性是不科学的。作者、期刊或者参考文献的影响力大只能说明文章在学术价值、内容质量或影响力方面较好,并不代表论文创新度高,因为论文的被引量和下载量受时间、作者和机构因素的影响,部分学者将论文影响力等同于创新力,这显然是片面的,论文创新力可能与影响力有相关关系,但并不等同于影响力。其次新颖性不等于创新性,新颖性只是作为论文创新的必要条件之一,不是充分条件,有的论文虽然研究对象很新,具有独创性,但是其应用价值和实践价值不高,或者其论证过程逻辑不合理,得出的结论并不有效,那么即便这种论文有很强的独创性,其创新价值也是不足的。正如李如森等[24]提出科技论文的创新点应分为主题中的创新点、技术背景中的创新点、技术方法中的创新点、论文结论中的创新点和总体创新点,并且指出论文创新成果应具有“独创性、新颖性、实用性”三个特点。这也体现了论文的创新不仅要在理论和方法上有创新,其创新成果还应是有效和有用的。

另外,采用文献计量学的方法评价论文创新性更多局限于定量地从论文的作者、机构、期刊以及参考文献等角度评价论文,而未能真正聚焦论文内容进行分析,虽然同行评议的定性评价方法在一定程度上弥补了这一缺陷,但其固有的弊端也影响了论文创新性的评价。如果能够运用自然语言处理技术和机器学习技术辅助论文创新性评价,通过语义分析构建语法规则,抽取创新表征词,提取论文创新点,构建创新知识库,并通过自动分类或聚类的方法识别创新主题类别,挖掘论文创新表达模式,则可以为论文创新评价提供坚实的技术保障和知识基础。自然语言处理技术在实现文献相似度的计算、主题的自动识别、关键词的抽取、主题的分类和聚类等方面已有成熟的应用。杨建林和钱玲飞[25]基于词频原则、逆文档频率原则以及共词分析的方法,构建了一套测量文档主题新颖度的计算公式,并采用实证法论证了这套公式的合理性和实用性。梁帅和高继平[26]以 F5000 收录的大量论文评审意见为文本分析对象,对优秀论文评审意见进行文本挖掘、关键词提取和内容分析,通过特征词频次和共现来分

chinaXiv:202304.00380v1

ChinaXiv合作期刊

析优秀论文的特征,文章指出优秀论文的特征关键词主要包括“创新特征”“价值特征”“研究内容”和“写作方式”四个方面。因此,判定论文是否具有创新价值不应该只局限在创新点的新颖性或独创性上,其创新成果的有用性和有效性,方法创新的科学性,都应该成为判断论文创新价值的评判要素。钱玲飞等[27]采用本体理论和技术,分别构造了学术创新力概念本体与学术创新力知识资源本体,将这些本体实例化,并引入 CNKI 期刊题录数据的高频关键词以丰富本体的知识,定义类属关系,构建了学术创新力本体,该研究成果为后续学术创新力自动测度研究提供了坚实的知识基础。贺婉莹[17]构建了用于创新力评价的机器学习模型,并利用图书情报领域核心期刊上的论文数据对多个模型进行实证分析,通过评估不同机器学习模型的表现得出了适合进行创新力评价的机器学习模型。可见,文本挖掘技术及机器学习方法在论文创新评价方面有相当的应用潜力和价值。

2.3 学术论文创新点识别与抽取

一篇学术论文可能有多种创新,也可能只有少量创新,但不管量的多少,只要是创新,哪怕是一个点、一句话都应识别出来并给予创新的肯定[28]。但是学术论文中创新的表述形式是多样的,创新点会以不同的形式出现在研究中的各个部分,因此对学术论文的创新点进行识别是非常有必要的。如上节所述,已有的研究成果大多是基于创新性评价,提出的对学术论文创新识别方法也大多从创新度评价角度出发。

曹小春[29]从编辑出版角度出发认为学术论文创新广义包括选题新、立论新、论证新、论据新,其方式有填补空白式、补充发展式、针锋相对式、破立变革式、引进式等,只要具备某一方面的特征,也就能被识别为学术论文的创新点。除了通过以上角度进行创新点的识别,还有研究的方式方法创新、研究思路新、研究设计新等[28]。

从学术论文的内容角度看,含有创新知识元就表示有创新性,而这个创新知识单元就是论文的创新点[30]。T. Heinze 等[31]认为可从研究中识别抽取提出新理论、发现新现象、提出和使用新方法、发明新仪器、从新角度整合现有理论等创新点。周露阳[16]从内容上的新论点、新论据逐步细化到新理论、新方法、新对策、新学科、新数据、新事实并继续深化提出了一套学术文献创新点的识别方法。也有学者认为可以通过参考文献的位置来定位与识别学术论文的创新点[32]。学者们提出的对学术论文的创新点识别多是宏观理论层面的,落到实际操作方面,还是会存在一定的差距。

对于学术论文创新点的抽取,一种方法是基于论文标题与数据库中旧的论文进行相似度排序,抽取出新的学术论文的创新点[33]。另一种方法是从论文中抽取出关键词,通过计算关键词频度,再与检索系统上随时间发展用户检索词的变化,抽取出创新关键词[25,34]。较全面的方法是用 Keygraph 算法[35]对论文的研究主题进行抽取,再将抽取出的研究主题与当前学科研究前沿进行相似度计算,抽取出创新研究主题[36]。也有研究人员考虑将论文上下文进行对比,通过对文本上下文进行新与旧的挖掘,识别抽取出技术或发明等创新点[37]。

国外对文献创新性的研究着眼点大都是从新闻事件、网页文件等文本新颖性出发,通过多种思路进行创新点的识别与抽取。M. Breja [6]将创新点分为事件级、句子级、信息级,认为新句子可以讨论新事件、也可以提供旧事件的新信息,本质上是通过句子级来进行创新识别的,通过句子的相似度以及新词数混合度量的方式进行句子新颖性排序以及创新点的挖掘。也有研究人员[5,38]从事件级角度出发对创新进行识别,通过文本相似度判断可能新颖的文本或事件[39]。

实际上以上所述大多数方法本质上是一致的,即与历史子集进行比较得到创新点。而这种方法存在缺点,即创建全面的历史子集难度和成本都非常高,从而可能导致标记为创新且可供训练使用的数据不足[40],识别效果变差。因此本文从另一个角度出发,选取中英文各两种情报学期刊作为样本,采用信息抽取和机器学习的方法,融合句子本身的特征识别领域创新内容,探讨学术论文的创新性特征,进而概括近年情报学领域的论文创新情况,探索创新点识别和抽取的新方法,助力学科领域科技论文创新性监测,从而促进科学研究创新。

3 研究设计

3.1 数据来源与文本预处理

本研究以论文写作的规范性和易获取性为考量指标,中文期刊论文以 2013 – 2018 年《情报科学》和《数据分析与知识发现》(原《现代图书情报技术》)为数据来源,英文期刊论文选取“Information Processing & Management”和“Journal of Informetrics”两期刊 2013 – 2018 年的论文作为数据来源,并对其中的会议纪要、主持人导语、发言稿、征文、选题等非研究论文进行了剔除,最终数据集分别为 2 487 篇和 1 050 篇。

ChinaXiv:202304.00380v1

在预处理部分,首先将所有论文格式转换为纯文本。虞沪生等[41]及 T. Dahl[42]的研究对论文创新点分布特征进行了总结,认为论文摘要、引言、结论等部分可以集中体现创新点。以此及统计结果为依据,提取可能出现创新句的论文摘要、引言、研究方法、结果、结论部分。之后对各部分内容进行了句子切分。

3.2 创新性特征引导词选取

论文创新点语言特征主要体现在引导词(特征词)和表达方式两方面[42-43]。针对科技文献的语言特征和体裁特征,利用基于规则的抽取方法可以准确识别论文中的"知识主张"[44]。对切分完成的句子使用Stanford CoreNLP 工具进行分词、词频统计和词性标注,结合随机抽取数据集人工标注的结果选取并确定了与创新紧密相关的语词,创新性特征引导词选择的主要依据来自"CSSCI 论文评测十分制评分标准"的"创新程度、完备程度、难易程度、成果价值"四要素,并精炼为"有用性、新颖性、有效性和科学性"。其中,新颖性是针对问题、方法和结果三大要素而言的,它是判断论文创新的核心要素:这种新颖性简单来说就是论文相较于其他已有文献的"不同",这种不同可以是对已有文献的"局部改进",也可以是与已有文献完全不同的"全新"[26]。有效性和科学性更多是对论文的研究方法的要求,有用性和有效性则是对研究结果的创新要求。初步确定创新性特征引导词后,将语词引入 Hownet 进行同义扩展,作为本文最终选取的创新性特征引导词。引导词包括但不限于标志性名词、形容词、少量动词和词组。本研究基于所选取的创新性特征引导词,同时参考张帆和乐小虬[44]研究的思路,通过 Stanford-parser 句法树解析,根据标注序列和结构构建引入创新性引导词的规则,对句子集进行抽取,构成创新句集。其中英文句子 19 088 个,中文句子 12 451 个。创新性特征引导词示例及对应例句如表 1 所示:

表 1 创新性特征引导词示例及对应例句

创新性特征引导词示例	例句
缺乏、突破	...以往研究多采用定性方法,进行理论归纳阐释,而缺乏定量客观研究评测,从而降低了理论成果的可实践性与可评估性。笔者试图将理论与实证相结合,以突破单一层面的偏颇性[45]。
解决	...该方法利用本体提供语义知识解决查询扩展过程中的语义偏差和歧义问题,结合用户查询意图进行初始查询扩展概念集的二次筛选,避免查询扩展过程中的检索主题偏移问题[46]。
验证、更好地	通过实验验证,该算法发现效果好,能够更好地获得热点话题[47]。
novel、approach introduced	Given these drawbacks, the principle of a novel data analytic citation prediction approach is introduced[48].
put forward	In this contribution we consider one particular node in a network, referred to as the ego. We combine Zipf lists and ego measures to put forward a conceptual framework for char acterizing this particular node[49].
no...exist、present the first…	The unfortunate truth is that no map of altruistic missions and causes exists; the landscape of altruistic activity is virtually unknown. In this paper, we present the first maps of altruistic mission space[50].

3.3 创新句句法分析

通过 Stanford-parser 对创新句进行依存句法分析,并构建规则抽取创新对象和创新主题。依存句法是由法国语言学家泰斯尼耶尔最早提出的,他认为句子中各个成分之间都存在着支配与从属关系。处于支配地位的词称为支配词或核心词,处于被支配地位的词称为从属词或修饰词[51]。依存句法分析采用词语对的二元关系形式体现句子中的词语之间的依存关系,通过定位语义标注类型为创新点句特征动词的谓词节点,可以进一步识别被其支配的主题词,而这些主题词即为可以揭示创新之处的核心主题词[8]。本研究根据依存句法关系的这一特性,给定创新对象和主题的抽取规则如下:①从依存关系对集合中识别出可以揭示创新点的核心词(即 ROOT 词);②从含有核心词的依存关系对集合中筛选直接宾语关系(即"dobj"标识符引导的关系),抽取其中的被支配成分作为揭示创新对象的语词;③扩展查找此关系对前后,与 ROOT 词和创新对象关联紧密(距离最近)的名词复合修饰关系对,将其作为本句的创新主题。对于中文句子,若存在"topic"关系对,则直接将此作为创新主题。对于不存在上述规则规定的关系对的句子,制定了基于句法树的补充规则:④对于英文句子,抽取标签为"JJ"(形容词)及其下位修饰标签为"NN"(常用名词)的关系对,并依据创新性引导词集对关系对进行筛选,将最终筛选得到的关系对中的被修饰部分作为创新对象;对于中文句子,由于标签更为丰富,所以扩展规则为抽取"ADJP"(形容词短语)及其下位修饰标签为"NP"(名词短语)的关系对,同样依据创新性引导词集对关系对进行筛选,将最终筛选得到的关系对中的被修饰部分作为创新对象;⑤抽取"IP"(简单从句)的最内层标签

chinaXiv:202304.00380v1

ChinaXiv合作期刊

"NP"作为创新主题。

3.4　创新句分类

李瑛和周立[52]在论文中对创新点进行了细化和具体化，将创新点类别归纳为{新发现、新方法、新技术、新观点、新理论、新思路、新工艺、新应用、新贡献、新设想}十个方面，并对这些创新点的表达内容和常用的特征词语进行了总结归纳。本研究以依存句法分析结果为基础，抽取创新句特征，加入了依存句法标签并向量化以提高识别的准确率，在借鉴李瑛和周立[52]的分类基础上，按照创新句描述的创新方面，将句子归为 4 大类 8 小类，分类及表达内容解释如表 2 所示：

表 2　创新句分类框架及类别内涵解释

序号	类别名称	类别主要表达内容
1	理论创新	发现新规律/联系
2		构建/改善/完善新模型
3		得到新理论
4		提出新方法/技术/思路
5		提出新对策/建议/应用
6	观点/概念创新	通过文献调研，采用某种理论/方法，对研究对象的概念进行归纳/完善/界定
7	研究方法创新	运用新方法、引入新数据
8	研究问题/对象创新	尚未解决/亟待解决的问题、鲜有某方面的研究、下一步研究计划、研究局限、研究视角的创新、研究热点、研究有待进一步深化

按照上述分类，采用 SVM 算法对人工标注的训练集进行训练和测试，并对余下的句子语料进行分类预测。结果的评价指标如表 3 和表 4 所示：

表 3　中文情报学期刊论文创新句分类结果评估指标

	precision	recall	f1-score
micro avg	0.84	0.69	0.75
macro avg	0.72	0.64	0.68

表 4　英文情报学期刊论文创新句分类结果评估指标

	precision	recall	f1-score
micro avg	0.84	0.77	0.81
macro avg	0.9	0.78	0.83

4　结果分析与讨论

4.1　创新句统计信息

对创新句集进行统计，得到平均每篇中文期刊论文 5 句左右，每篇英文期刊论文 18 句左右。各部分创新句分布比例见图 1、图 2。

对比图 1 和图 2，可以发现创新句在中外情报学论文中分布情况较为一致，英文论文的创新句分布更为均衡。有 30% 以上的创新句出自引言部分，一方面是由于引言部分一般是论文的必备项，本身长度较摘要长，相对于摘要对创新点表达的简洁，引言作为论文的第一部分需要对创新的表达进行拓展和丰富，除了适当介绍创新性内容外，还包括创新的缘起、目的、手段，创新点的呈现方式以叙述式表达为主[52]。由于作者"可以自定义论文主体部分的编写格式"[53]，使得"研究结果"并不是论文的必要部分，其呈现方式十分多样，因此这一部分创新句比例不高。但所选英文情报学期刊论文中研究结果部分的比例仍高于摘要，说明英文情报学期刊论文写作相对规范。而结论的创新点揭示用语较少，主要是表述创新的价值、作用和意义，对创新性本身的表述通常是隐性和间接的。总体来看，中英文情报学期刊论文的创新句分布比例差距不大，英文期刊论文的篇平均数更高，对创新点的表达更多。

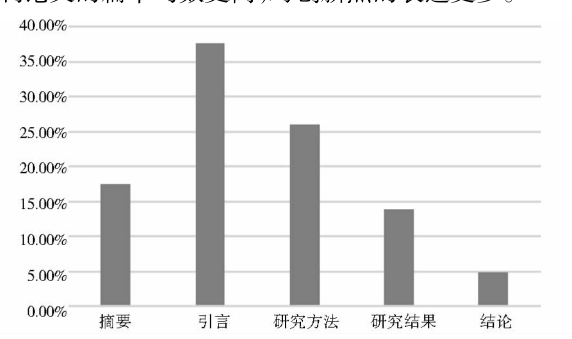


图 1　中文期刊论文中创新句在各部分的分布比例

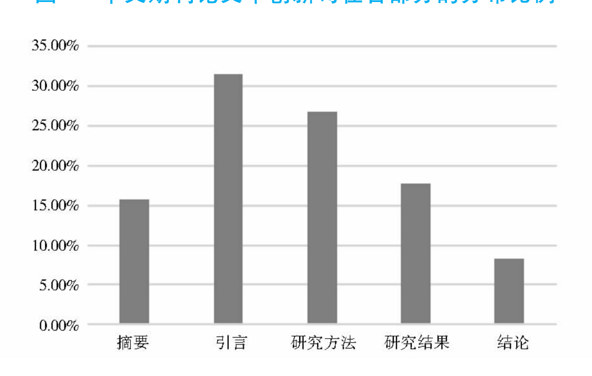


图 2　英文期刊论文中创新句在各部分的分布比例

4.2　论文创新对象特征分析

创新对象的词频及对应序号见图 3、图 4。可以看到，中英文情报学论文的创新对象频率排序都基本符合齐普夫定律，即排序靠前的创新对象占据了创新对象的绝大部分，而排序靠后的创新对象则数量非常稀少。相比之下，英文情报学期刊论文的图像更为"陡峭"，反映了英文情报学期刊论文的创新对象更为集中。且由于英文论文的创新对象数量更多，其长尾效应也更为明显。

chinaXiv:202304.00380v1

ChinaXiv合作期刊

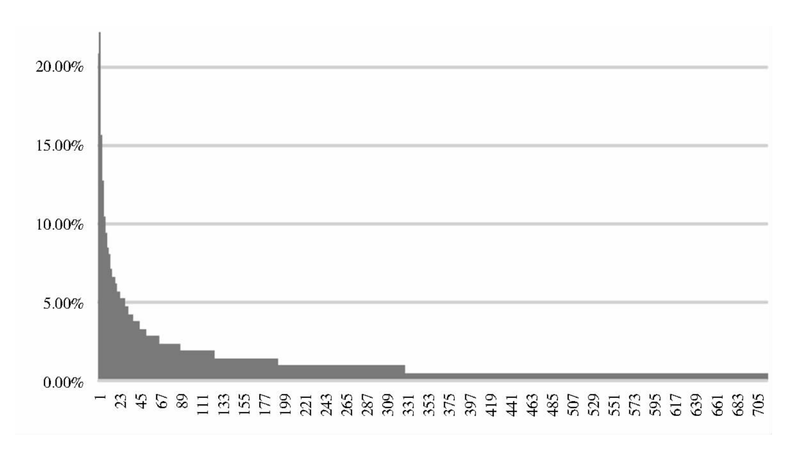


图 3　中文期刊论文中创新对象频率分布

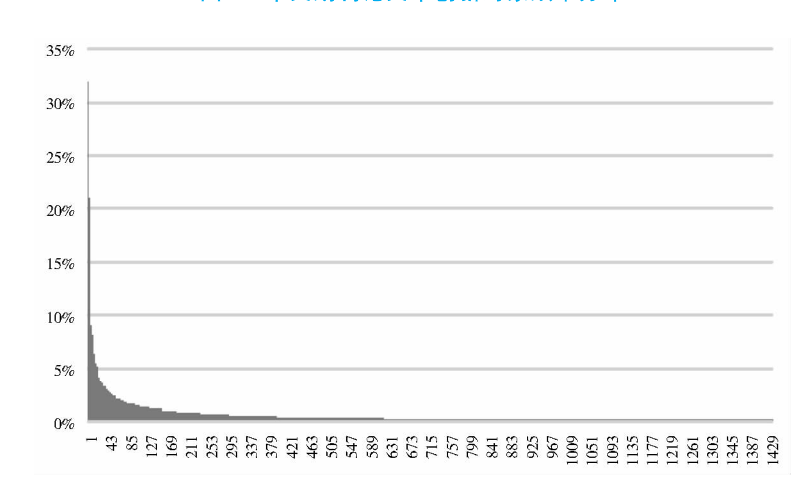


图 4　英文期刊论文中创新对象频率分布

chinaXiv:202304.00380v1

对创新对象进行具体探究,可以看到“方法”上的创新在中英文情报学期刊论文中都占了非常大的比重,说明方法的创新是近年情报学领域的重点研究方向。而从相对频率来看,“方法”创新在英文论文中占比更高,“method”和“approach”的频率之和超过一半,说明方法创新在英文情报学期刊论文中占大部分。而研究方法的“方法”创新,即方法论(methodology)的创新则占 4% 左右,比例较小。相比之下,中文论文中的方法创新比例虽然最高,但其中包含 1% 左右的研究方法创新,高层次方法创新的比例仍然较低。魏瑞斌[54]通过对国内共词分析研究的部分成果的进行分析,同样发现方法研究的整体创新的论文数量很少。他认为研究方法创新需要从方法的原理层面有所突破,或者是对方法的某些流程进行改进。这需要研究者对研究方法有非常深入的理解,并能够提出自己的解决方案,因而其创新难度较大。中文论文各创新对象频率差距相对较小,除去一些泛用词,可以看到,除方法之外,按照频次顺序,数据、问题、算法、模型、技术、理论是中文情报学期刊论文进行创新的重点对象,模型、结果、框架、算法、问题、数据、测量是英文情报学期刊论文进行创新的重点对象。二者重合的部分仍比较多(如数据、问题、模型、算法),这些也是近年情报学领域的主要创新方向。相比之下,英文情报学期刊论文更侧重细节上的创新,在理论方面创新较少,但更多地关注了“框架(framework)”的创新(见图 5、图 6、表 5)。这与刘齐进等[9]得到的结论类似。刘齐进等对 1951－2012 年间《ACM 美国计算机学会全文数据库》中收录的“计算机”学科的 21 万多篇英文文献进行分析,发现该领域大部分创新集中在方法的创新(如 approach, method, way),以及具体应用的创新(如 algorithm, model, application),关于理论的创新(如 idea)则相对较少[9]。

对比两中文情报学期刊(见表 6),可以发现论文创新对象呈现明显的不同,主要与期刊的栏目设置及定位有关。《情报科学》设有理论研究和业务研究的栏目,因此期刊的理论、技术、方法创新相对占比较高。而《数据分析与知识发现》聚焦各行各业中以大数据为基础、依靠复杂挖掘分析、进行知识发现与预测、支持决策分析和政策制定的研究与应用,致力于提供理

ChinaXiv合作期刊
chinaXiv:202304.00380v1

论指导、技术支持和最佳实践。依靠并融汇计算机科学、科学计量学、社会计量学、网络计量学、数据科学、管理科学、预测分析、循证政策分析等领域，帮助人们从数据发现知识、从知识提炼智慧（洞察力）、从知识和智慧推演并设计解决方案[55]，这样的倾向使其论文中算法、实验等方面的创新比较突出。

两本英文情报学期刊的高频创新对象相对比较一致（见表 7），不同的部分同样主要与期刊的定位有关。除方法和模型外，*Information Processing & Management* 期刊论文的创新对象比较宽泛、全面，算法、框架、研究问题、实验、特征、技术等都有包括，期刊本身定位也在计算机和信息科学的交叉领域。*Journal of Informetrics* 的论文以信息科学定量的研究为主，因此其论文的创新对象特点比较鲜明，指标、数据、利用的创新比较突出。

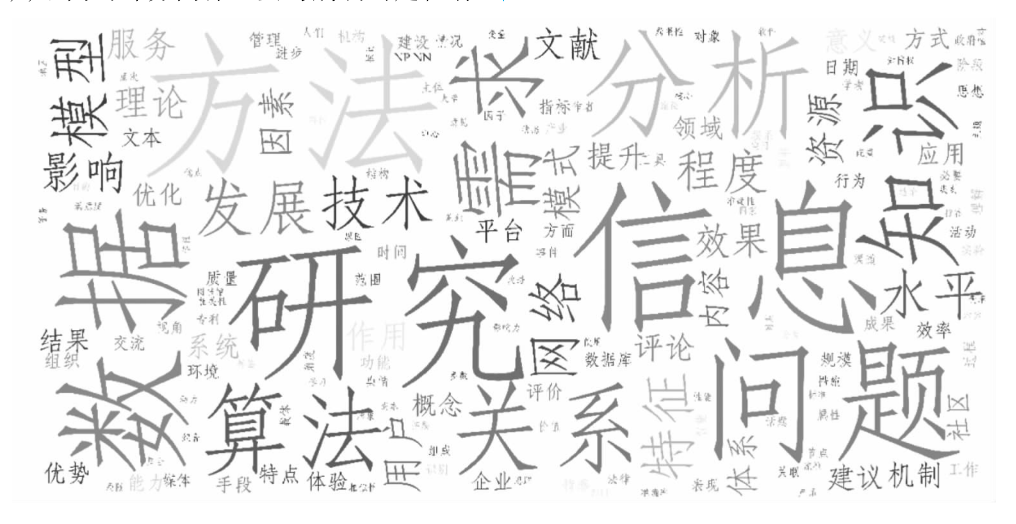


图 5　中文期刊论文创新对象频率 Top200 词云

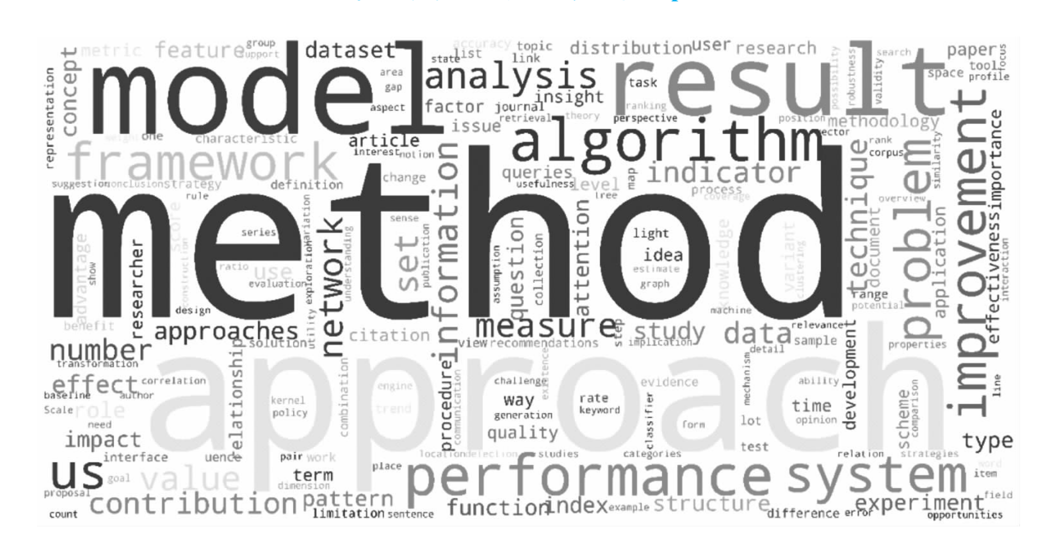


图 6　英文期刊论文创新对象频率 Top200 词云

表 5　中英文期刊论文创新对象重合部分（高频）

中文论文创新对象	频率占比	英文论文创新对象	频率占比
方法	22.28%	method	32.03%
数据	15.64%	approach	21.07%
问题	12.80%	model	16.27%
模型/算法	9.48%	algorithm	8.73%
		problem	5.82%
		data	5.48%

4.3　论文创新主题特征分析

对创新主题进行共现分析，得到图 7 和图 8 所示的共现网络。边的粗细代表边的权重，节点的颜色表示模块化分类的情况。可以很明显地看出，在节点数相差不大的情况下，两图边数差距非常大，中文情报学主题共现网络的图密度为 0.029，英文情报学主题共现网络的图密度为 0.649。情报学领域的英文论文各主题语词共现频繁，交集更多，联系紧密，并以“方法”

表 6　中文情报学期刊论文创新对象对比

《情报科学》创新对象	频率占比	《数据分析与知识发现》创新对象	频率占比
数据	17.34%	方法	41.77%
问题	16.85%	算法	25.42%
方法	15.63%	特征	19.97%
技术	9.63%	问题	18.16%
理论	8.99%	模型	16.34%
模型	8.35%	数据	12.71%
特征	5.78%	效果	12.71%
结果	5.14%	技术	10.90%
算法	5.14%	模式	9.08%
应用	5.14%	实验	7.26%

表 7　英文情报学期刊论文创新对象对比

"Information Processing & Management" 创新对象	频率占比	"Journal of Informetrics" 创新对象	频率占比
method	36.50%	method	25.04%
approach	24.30%	approach	17.00%
model	20.60%	model	10.00%
performance	12.90%	index	9.00%
results	13.30%	indicators	8.00%
algorithm	12.00%	data	7.20%
framework	11.00%	results	7.20%
system	9.00%	framework	5.80%
improvement	8.00%	methodology	5.40%
problem	75.00%	technique	5.40%
experiment	6.00%	performance	4.90%
feature	5.00%	system	4.50%
measure	4.00%	measure	4.02%
technique	3.00%	use	4.02%

chinaXiv:202304.00380v1

"数据""模型"等为核心构成非常密集的网络。相比之下,中文情报学研究主题相对分散和独立。近年英文情报学的创新主题大致可以分为:文献计量类(citation、indicator、journal、research evaluation、collaboration…);文本挖掘类(classification、text、topic、detection…);信息系统相关(system、search、framework、recommendation…)以及与自然语言结合的机器学习类(algorithm、cluster、language、word、training、sentiment…)。其中文献计量大类包含的主题与刘智锋等[56]对"Journal of Informetrics"的研究结果较为一致。刘智锋等选取 2007 – 2017 年"Journal of Informetrics"论文作为标注数据源,构建语义功能标注数据集并进行分析,发现信息计量学领域主要的研究主题包含有计量指标、科学评价及排名、科研合作、引用分析[56],反映了这些主题是这个领域核心的创新研究主题,且相对稳定。近年中文情报学的创新主题大致可以分为:图书馆相关类(数字图书馆、高校图书馆、图书馆服务、资源……);算法类(遗传算法、优化算法、神经网络……);社交媒体类(社交网站、话题……);社区研究类(科研/学术/健康社区、成员、影响力……);企业知识管理类(知识转移、个体能力、创新……);本体类以及文献计量类 7 个大类。同样由于联系过于紧密,英文主题难以像中文一样进行更细的区分,但也能看出,在高频主题中英文情报学期刊论文以自然语言处理和文献计量相关内容为主,主题较集中,而中文情报学期刊论文主题更加分散和多样。这虽然与期刊选择有一定关系,但也能一定程度上反映中文情报学期刊收录论文的综合性和英文情报学期刊的专业性以及研究主题的不同特点。

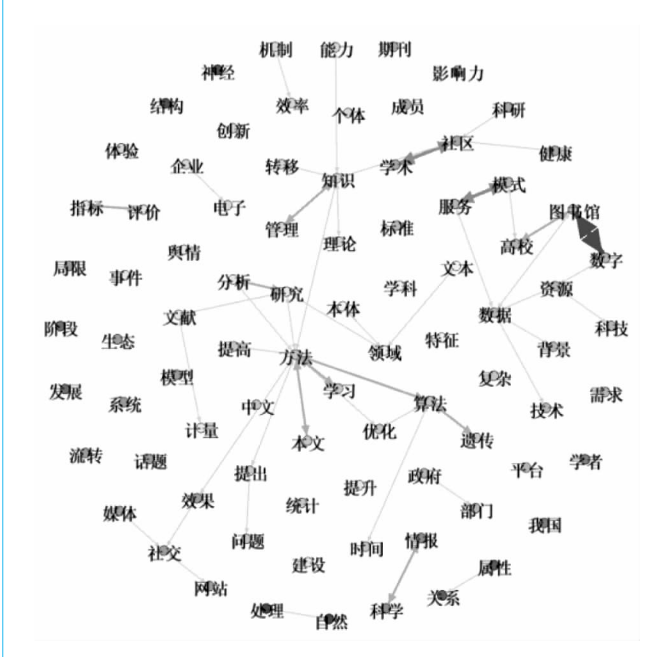


图 7　中文期刊论文创新主题共现网络

4.4　论文创新类别分布

论文创新句的类别分布如图 9 和图 10 所示,并将各类创新的核心表达范式归纳为表 7 和表 8。中文论文中各类创新句的分布呈现不均衡的状态。第 6 类和第 7 类即观点/概念创新和研究方法创新较少,而第 4 类和第 8 类即提出新方法/技术/思路和研究问题/对象创新占比则较高。相比之下,英文论文创新点分布更加不均衡,将近 80% 的创新句集中分布在第 1 类(发现新规律/联系)、第 4 类和第 8 类的创新类别上,第 2 类、第 3 类、第 5 类和第 6 类则较少。英文和中文论文的创新类别都集中分布在第 4 类和第 8 类,即提出新方法/技术/思路和研究问题/对象创新上。这与"论文创新对象特征分析"一节的结果形成呼应,即研究问题和针对某一具体研究对象提出新方法是近年中

ChinaXiv合作期刊

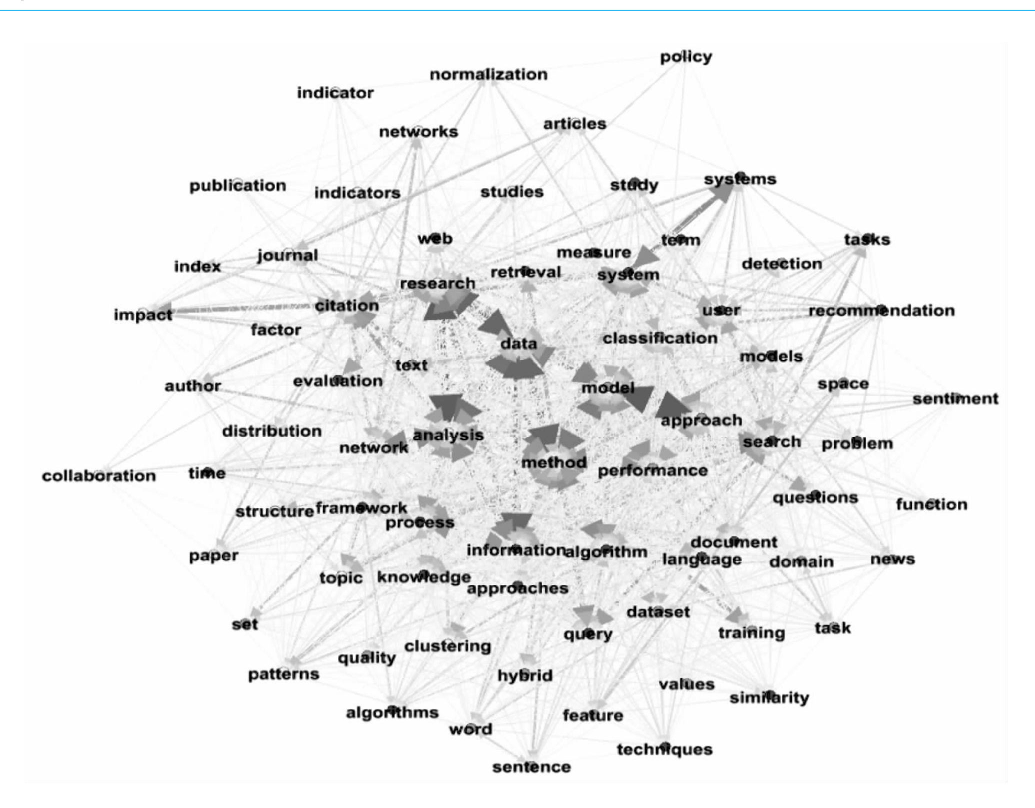


图 8　英文期刊论文创新主题共现网络

ChinaXiv:202304.00380v1

外情报学期刊论文的主要创新方向。从两图中可以看出,中英论文创新点分布的最大差异体现在第 2 类创新(构建/改善/完善新模型)、第 5 类创新(提出新对策/建议/应用)和第 7 类创新(运用新方法、引入新数据)的占比上。中文论文的第 2 类和第 5 类创新占比相对较高,反映了中文论文中综述类和定性评价类占比相对较高,中文的综述类论文通过梳理国内外相关主题的研究理论成果和实践进展来提出对某一产品、某一行业、学科或领域建设发展的对策和应用,评价类论文多综合运用专家访谈法、问卷调查法和层次分析法构建评价指标体系,少数会结合数理统计学的知识或自然语言处理技术对模型或方法的信度和效度进行实证分析检验。这也反映了我国情报学期刊论文在研究方法上创新不足,运用其他学科的研究方法、引入新数据不是主流。而英文论文中第 7 类比重相对较高,说明英文情报学期刊刊载的研究不仅提出的具体新方法多,运用新研究方法的也更多,引入新数据的情况也相对较多,在定量实证方向上比较占优势,侧重于研究成果的应用性和实践性。这也一定程度上反映了中西方学者在论文写作方面思维的差异,纪蓉琴[56]认为“西方文化中数理逻辑的发达以及西方人的逻辑型理性思维模式决定了英语文章在陈述论证时,倾向于较多地引用事实,尤其是调查数据和数据事实。相比之下,汉语文章则较少运用数据和实验结果来论证观点”。第 3 类(得到新理论)和第 6 类(观点/概念创新)在中英文论文中占比都很低,进一步反映了近年情报学理论创新推动缓慢的事实。魏瑞斌和刘宇[57]通过对 1996 – 2012 年期间的 434 篇情报学博士论文的标题进行文本分析,也认为国内情报学论文当中,应用研究、实证研究的成果较多,而纯理论的研究成果较少,这也是由于理论创新是论文创新中最有价值也是难度最大的创新方式。

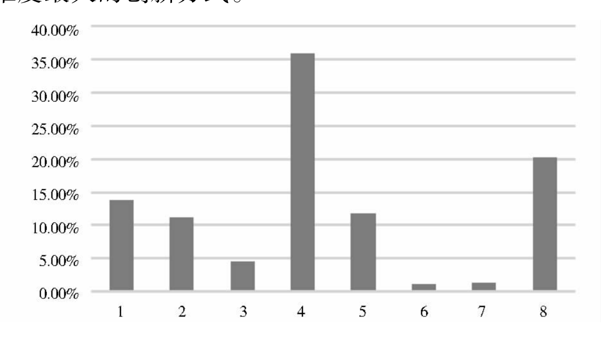


图 9　中文期刊论文中各类句子分布比例

4.5　论文创新表达范式

通过对基于机器学习训练出的分类集合的内容进行定性的分析,本文对中英 8 种类别的论文创新表达范式进行了归纳总结,得到如表 8 和表 9 所示的结果。可以看出无论是哪种类别的创新,都是建立在对现有文献的回顾和评估的基础上的,通过文献回顾了解研究的理论基础,理出研究问题的发展脉络,了解之前的

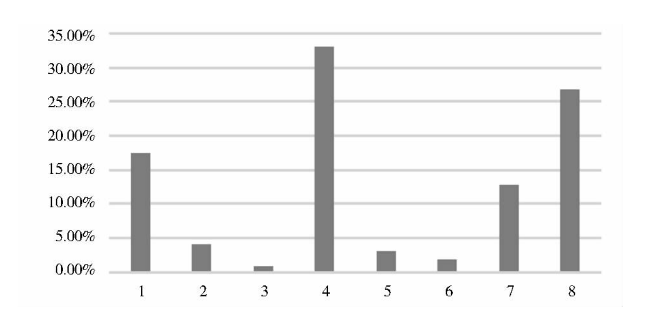


图10 英文期刊论文中各类句子分布比例

研究已经解决了什么问题，还未解决什么问题，有何不足之处，从而为自身的研究创新提供新的方向和思路。其次，创新方法需要真实有效，即论据经得起考证，数据真实有效，提出的方法能够通过实验的检验。最后，创新结果需要有价值，这从特征词提升、优化、改善、有效的、理论价值、实践价值等可以看出一二。

表8 中文情报学期刊论文各类创新核心表达范式

创新类别	核心表达范式
发现新规律/联系	从...理论出发/借鉴…理论，运用…方法进行/展开分析，发现…(规律)，提升了…水平/效果
构建/改善/完善新模型	研究基于…构建模型解决…问题/引入…改进现有模型，通过实验证明/表明…预测效果/性能/精度得到优化/提升
观点/概念创新	现有研究缺乏对…的界定，通过文献调研，在…分析的基础上，将...定义为/认为…
提出新方法/技术/思路	针对…问题，为改善/优化/填补现有研究，提出了…方法/算法，实验证明该方法能有效…，为…提供新思路/途径/参考
研究方法创新	利用/使用/应用…方法对…分析，解决…问题/对象，得到更好的…效果
研究问题/对象创新	(同类主题或研究对象的研究中)存在缺陷/不足/问题，需要进一步/建立/考虑…，或者从…视角出发/提供了…的新视角
提出新对策/建议/应用	在...基础上提出...建议，将…应用于以…得到新的效果
得到新理论	针对…研究中存在的缺陷/鲜有…方面的研究/针对…新的现象进行研究，或者将…的技术、方法、理论应用到新的对象(领域、学科、事件)上

5 研究结论

从论文中识别出创新特征与研究新进展对于领域内的科学研究具有重要意义。本研究从句子级创新性识别出发，利用句子结构和句法特征，采用信息抽取和机器学习的方法分析创新的类别、对象及主题，从实证研究角度揭示情报学近年的创新特征与进展。研究得出以下结论：首先，从创新的表达来看，中外情报学论文创新句的分布情况基本一致，其中引言是创新点集中表述的部分，但是英文期刊论文在行文方面更加规范，创新的表达更丰富；中英文创新点的表达都较为规范，有一定规律可循。其次，从创新性特征来看，分析创新对象特征和类别分布发现，创新对象频率分布符合齐普夫定律，即高频创新对象占据了创新对象的绝大部分数量，具体方法的创新是近年情报学领域研究的主要方向，数据、问题、模型、算法等的创新也是研究人员们比较关注的方面，而在研究方法上创新不足。相较而言，英文情报学期刊论文更侧重细节上的创新，在理论方面创新更少，但较多地关注了“框架”的创新；中英文情报学期刊论文的创新性特点都反映了应用研究、实证研究的成果较多，而理论创新推动缓慢，这也是由于理论创新是论文创新中最有价值也是难度最大的创新方式。对论文创新主题特征分析发现，英文情报学期刊论文以自然语言处理和文献计量相关内容为主，主题较集中，而中文情报学期刊论文主题具有多样和分散的特点，一定程度上反映了中文情报学期刊收录论文的综合性和英文情报学期刊的专业性以及研究主题有差异的特点。

表9 英文情报学期刊论文各类创新核心表达范式

创新类别	核心表达范式
发现新规律/联系	it was /we found that…(v) more/higher/better than…
构建/改善/完善新模型	We propose/introduce/develop new model(s) for/that/based on… The proposed model can…
观点/概念创新	…is defined as follows… We defined…to…
提出新方法/技术/思路	In this paper, we deploy/propose an approach that can/to … …approach is proposed/ …has been solved following a …approach. the proposed approach has improved/outperforms…
研究方法创新	our dataset covers… we use …data about…/apply…methodology introduced by …
研究问题/对象创新	the question…ask for… It also leads to new questions about … Our study aims to address the following questions …
提出新对策/建议/应用	the suggestions we provided will…
得到新理论	By/based on…, we demonstrate that….

基于以上研究结论，情报学领域的研究者不仅应该关注具体层面的方法、问题、技术创新，还应该更多地尝试难度更大的、更高层次的研究方法和理论创新，注重学科交叉，寻找新的突破点，促进情报学科理论发展由量变到质变，为情报学向更成熟的层面发展奠定理论基础。对情报学研究者来说，以上结论还可以为日后研究的创新方向和论文写作中的创新点表达提供参考。本文的初步尝试还表明，通过信息抽取和句法

chinaXiv:202004.00038v1

ChinaXiv合作期刊

分析对论文创新点进行分析具有一定的可行性和价值。这表现为不仅可以借助这种方法进行更细粒度的论文创新性分析，进而对整个领域的研究创新进展有所把握，实现监测的目的，从而促进科学研究创新，还可以为创新性的内容评估提供方法参考，成为现有创新性评价体系的补充。

然而，本研究也存在一定的局限性：研究利用句子结构和句法特征对单一句子进行创新性识别，没有深入考虑句子上下文之间逻辑层面的语义联系，这可能会对研究结果造成一定的影响。另外，研究只抽取了国内外各两种期刊的论文，样本的覆盖面还不够全面；句子抽取结果受句法分析影响，对否定词等考虑不周，相关抽取规则还待完善。针对以上提到的研究局限性，在未来的研究中将进一步合理优化创新的分类结构，改进对论文创新句抽取算法，进一步提高抽取的全面性，同时通过考虑论文上下句的关系并将论文发表的时间轴添加进去；另外，选择领域内或者更多学科的期刊论文进行实证研究，并加入时间线的对比，更加全面地探究论文创新性特征，助力更好地实现细粒度知识组织和检索。

chinaXiv:202004.00038v1

参考文献：

[1] TAN R. On the declaration of novelty in scientific journal articles[EB/OL]. [2019-10-30]. https://www.philstar.com/business/science-and-environment/2014/04/24/1315251/declaration-novelty-scientific-journal-articles.

[2] 徐书荣. 科技期刊编辑对提升论文创新性的作用[J]. 中国科技期刊研究, 2014,25(6): 761-764.

[3] GABRILOVICH E, DUMAIS S T, HORVITZ E, et al. Newsjunkie: providing personalized newsfeeds via analysis of information novelty[C]//Proceedings of the 13th international conference on World Wide Web. New York: ACM, 2004: 482-490.

[4] OBEID N, RAO R B K N. On integrating event definition and event detection[J]. Knowledge and information systems, 2010, 22(2):129-158.

[5] TSAI F S, CHAN K L. Redundancy and novelty mining in the business blogosphere[J]. The learning organization, 2010, 17(6): 490-499.

[6] BREJA M. A novel approach for novelty detection of web documents[J]. International journal of computer science and information technologies, 2015, 6(5): 4257-4262.

[7] LI X, CROFT W B. Novelty detection based on sentence level patterns[C]//Proceedings of the 14th ACM international conference on Information and knowledge management. New York: ACM, 2005: 744-751.

[8] 张帆，乐小虬. 领域科技文献创新点句中主题属性实例识别方法研究[J]. 数据分析与知识发现, 2015, 31(5):15-23.

[9] 温浩. 科技文摘创新点语义识别与分类方法研究[J]. 情报学报, 2019, 38(03):27-34.

[10] 刘齐进，程齐凯，陆伟. 学术文献创新修饰词描述对象分析[C]//南京：第九届全国情报学博士生学术论坛. 2019.

[11] Merriam-Webster, Incorporated. Merriam-Webster Online Dictionary. [EB/OL]. [2019-10-30]. http://www.merriam-webster.com.

[12] NG K W, TSAI F S, CHEN L C L, et al. Novelty detection for text documents using named entity recognition[C]// International Conference on Information. Piscataway: IEEE, 2008.

[13] ALLAN J, WADE C, BOLIVAR A. Retrieval and novelty detection at the sentence level[C]//Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. New York: ACM, 2003: 314-321.

[14] UZZI B, MUKHERJEE S, STRINGER M J, et al. Atypical combinations and scientific impact[J]. Science, 2013, 342(6157): 468-472.

[15] ARTHUR W B. The nature of technology: what it is and how it evolves[M]. New York: Simon and schuster, 2009.

[16] 周露阳. 论审评学术论文创新因素的指标体系[J]. 编辑学报, 2006(01):68-70.

[17] 贺婉莹. 基于机器学习的论文学术创新力评价研究[D]. 南京：南京大学, 2019.

[18] 刘丽萍，刘春丽. 开放同行评议利弊分析与建议[J]. 中国科技期刊研究, 2017, 28(05):389-395.

[19] 王晓慧，王康. 基于期刊论文的学者学术影响力评价及评价指标关系研究——以竞争情报研究领域为例[J]. 情报科学, 2018, 36(02):63-66, 87.

[20] LEE Y, WALSH J, WANG J, et al. Creativity in scientific teams: unpacking novelty and impact[J]. Research policy, 2015, 44(3): 684-697.

[21] WU L, WANG D, EVANS J A, et al. Large teams develop and small teams disrupt science and technology[J]. Nature, 2019, 566(7744): 378-382.

[22] 宋歌. 科研成果创新力指标 S 指数的设计与实证[J]. 图书情报工作, 2016, 60(05):77-86, 124.

[23] 叶继元. 学术期刊的质量与创新评价[J]. 浙江大学学报(人文社会科学版), 2013, 43(02):108-117.

[24] 李如森，彭彩红，赵福荣. 科技论文创新性判断方法[J]. 鞍山钢铁学院学报, 2001(03):234-236.

[25] 杨建林，钱玲飞. 基于关键词对逆文档频率的主题新颖度度量方法[J]. 情报理论与实践, 2013, 36(03):99-102.

[26] 梁帅，高继平. 基于 F5000 论文评审意见的优秀论文特征识别[J]. 科学学研究, 2017, 35(03):331-337.

[27] 钱玲飞，张吉玉，汪荣，等. 基于领域知识的学术创新力测度本

体构建研究[J]. 现代情报,2019,39(05):30－37.
[28] 崔京艳,邓媛,李春梅,等. 中医学科技论文创新性和科学性的识别与评价[J]. 中国医药导报,2018,15(05):172－176.
[29] 曹小春. 学术期刊审稿专家的职责[J]. 编辑之友,2006(06):62－63.
[30] 索传军. 知识转移视角下的学术论文老化与创新研究[J]. 图书情报工作,2014,58(05):5－12.
[31] HEINZE T, SHAPIRA P, SENKER J, et al. Identifying creative research accomplishments: Methodology and results for nanotechnology and human genetics[J]. Scientometrics, 2007, 70(1): 125－152.
[32] 朱大明. 参考文献的主要作用与学术论文的创新性评审[J]. 编辑学报,2004(02):91－92.
[33] CANNON D C, YANG J J, MATHIAS S L, et al. TIN-X: target importance and novelty explorer[J]. Bioinformatics, 2017, 33(16): 2601－2603.
[34] 沈阳. 一种基于关键词的创新度评价方法[J]. 情报理论与实践,2007(01):125－127.
[35] OHSAWA Y, BENSON N E, YACHIDA M. KeyGraph: automatic indexing by co -occurrence graph based on building construction metaphor[C]// Proceedings of IEEE international forum on research and technology advances in digital libraries. Washington: IEEE,1998: 12－18.
[36] 杨京,王芳,白如江. 一种基于研究主题对比的单篇学术论文创新力评价方法[J]. 图书情报工作,2018,62(17):75－83.
[37] THORLEUCHTER D. Finding new technological ideas and inventions with text mining and technique philosophy[M]//Data analysis, machine learning and applications. Springer, Berlin, Heidelberg, 2008: 413－420.
[38] TANG W, TSAI F S, CHEN L. Blended metrics for novel sentence mining[J]. Expert systems with applications, 2010, 37(7): 5172－5177.
[39] FU X , CH'NG E , AICKELIN U . An improved system for sentence-level novelty detection in textual streams[C]// 3rd International Conference on Smart Sustainable City and Big Data (ICSSC). London:IET, 2015.
[40] AMORIM M, BORTOLOTI F D, CIARELLI P M, et al. Novelty detection in social media by fusing text and image into a single structure[J]. IEEE access, 2019, 7: 132786－132802.
[41] 虞沪生,张瑞清,阎为民. 科技论文创新性的审读[J]. 编辑学报,2006(05):333－334.
[42] DAHL T. The linguistic representation of rhetorical function: a study of how economists present their knowledge claims[J]. Written communication, 2009, 26(4): 370－391.
[43] PARKINSON J. The discussion section as argument: the language used to prove knowledge claims[J]. English for specific purposes, 2011, 30(3): 164－175.
[44] 张帆,乐小虬. 面向领域科技文献的句子级创新点抽取研究[J]. 现代图书情报技术,2014(09):15－21.
[45] 李璐. 信息资源产业与文化产业融合的实证分析——基于中国上市公司1997年－2012年数据[J]. 情报科学,2016,34(03):122－126.
[46] 李爱明. 基于本体和用户查询意图的查询扩展方法研究[J]. 情报科学,2015,33(05):68－71.
[47] 魏德志,陈福集,林丽娜. 一种基于时间序列的热点话题发现模型和算法[J]. 情报科学,2017,35(10):142－146.
[48] CAO X, CHEN Y, LIU K J, et al. A data analytic approach to quantifying scientific impact[J]. Journal of informetrics, 2016, 10(2): 471－484.
[49] ROUSSEAU R, ZHAO S X. A general conceptual framework for characterizing the ego in a network[J]. Journal of informetrics, 2015, 9(1): 145－149.
[50] KLAVANS R, BOYACK K W. Mapping altruism[J]. Journal of informetrics, 2014, 8(2):431－447.
[51] 郑捷. NLP 汉语自然语言处理[M]. 北京:电子工业出版社,2017;283－287.
[52] 李瑛,周立. 科技期刊论文创新点合理呈现的价值及理想模式[J]. 中国科技期刊研究,2018,29(10):993－999.
[53] GB 7713－1987. 科学技术报告、学位论文和学术论文的编写格式[S]. 北京:中国标准出版社,1987.
[54] 魏瑞斌. 基于内容分析的国内图书情报学研究方法创新研究——以共词分析方法为例[J]. 图书情报工作,2016,60(24):107－114.
[55]《数据分析与知识发现》编辑部.《数据分析与知识发现》介绍[EB/OL]. [2019－11－02]. http://manu44.magtech.com.cn/Jwk_infotech_wk3/CN/column/column291.shtml.
[56] 刘智锋,李信,程齐凯,等. 学术文本关键词语义功能数据集构建与分析——以 Journal of Informetrics 为例[J]. 图书馆论坛,2019,39(07):64－74.
[56] 纪蓉琴. 中英思维模式差异对英汉学术论文写作的影响[J]. 华东交通大学学报,2006(06):134－137.
[57] 魏瑞斌,刘宇. 基于标题文本分析的博士论文选题创新性研究——以国内情报学博士论文为例[J]. 情报杂志,2017,36(07):122－127.

chinaXiv:202304.00380v1

作者贡献说明：

曹树金：提出研究问题、思路和框架，论文修改；
闫欣阳：设计实验方案，负责实验分析和数据处理，论文撰写；
张倩：负责实验分析和数据处理，论文撰写；
卓伊玲：负责实验分析和数据处理，论文撰写。

ChinaXiv合作期刊

Research on Characteristics of Innovation in Chinese and International Academic Literature of Information Science

Cao Shujin Yan Xinyang Zhang Qian Zhuo Yiling
School of Information Management, Sun Yat-sen University, Guangzhou 510006

Abstract: [Purpose/significance] This paper comprehensively uses qualitative and quantitative methods to analyze and compare the innovative features and expression paradigms of Chinese and foreign information science papers in recent years, thus revealing the innovative characteristics of papers in the field of information science, discovering the knowledge relationship within the innovative sentences, and conducting a more fine-grained analysis of the innovation of the papers, which provides conditions for the discovery and utilization of innovations in the research field, and at the same time, enriches the ways of innovation monitoring of research papers and promotes scientific research innovation. [Method/process] Starting from the sentence-level innovative identification, two kinds of information science journals in Chinese and English are selected as samples, and the methods of NLP and machine learning are used to the process of information extraction and classification. This paper extended the extraction of innovative sentences from the existing abstract to the full text, moreover made full use of sentence structure and syntactic features to identify innovative content in the field, and explored the characteristics of Chinese and foreign information science papers in terms of innovation objects, themes, categories, etc., meanwhile made a comparative analysis. Finally, through qualitative analysis of the automatic classification of document collections, the paper summarized the expression paradigms of the innovation of Chinese and foreign papers. [Result/conclusion] From the expression of innovation points, the distribution of innovative sentences in Chinese and foreign information science papers is basically the same. The expression of innovation in foreign papers is richer. In terms of the characteristics of innovation, the innovation topics of foreign information science journals are concentrated, while the Chinese ones are diversified and scattered. The innovation of specific methods is a hot spot in the field of information science in recent years. The innovation characteristics of both Chinese and foreign information science journal papers reflect the results of applied research and empirical research are richer, while the trend of theoretical innovation is slow.

Keywords: characteristics of innovation academic literature dependency parsing classification of sentences

chinaXiv:202304.00380v1

下 期 要 目

- □ 专题:政府开放数据用户行为与服务研究 （段尧清教授组织）
- □ 期刊论文视域下我国图情学科的研究模式探究 （黄国彬 郑霞 王琼）
- □ 基于组合赋权 - TOPSIS 法的高校图书馆数字资源服务绩效评价 （陈英）
- □ 高校图书馆空间认知实证研究——以利益相关者为视角 （伍玉伟 洪芳林）
- □ 免费与付费在线问答社区用户参与行为的比较研究 （齐云飞 赵宇翔 刘周颖等）
- □ 国外图书馆展览服务研究与实践及借鉴 （王峥）